Available online at www.sciencedirect.com

ScienceDirect

Journal of Multivariate Analysis 98 (2007) 1955–1968

Journal of Multivariate Analysis

www.elsevier.com/locate/jmva

Deconvolution from panel data with unknown error distribution

Michael H. Neumann

Friedrich-Schiller-Universität Jena, Institut für Stochastik, Ernst-Abbe-Platz 2, 07743 Jena, Germany

Received 25 January 2006
Available online 9 November 2006

Abstract

We devise a new method of estimating a distribution in a deconvolution model with panel data and an unknown distribution of the additive errors. We prove strong consistency under a minimal condition concerning the zero sets of the involved characteristic functions.

AMS 2000 subject classification: primary 62G05; secondary 62G20

Keywords: Minimum distance; Nonparametric deconvolution; Strong consistency; Panel data; Unknown error distribution

1. Introduction

Nonparametric deconvolution is one of the standard problems in statistics. It appears when a variable of interest can only be observed with some contamination which is modelled as an independent additive error.

There already exists a considerable amount of literature for the case that the distribution of the unobserved errors is known. The most frequently used approach to estimate the density of interest is the kernel method, which amounts to a damped or truncated division of the empirical characteristic function of the observations by the characteristic function of the errors. Consistency and rates of convergence together with their optimality are proved in Carroll and Hall [4], Devroye [7], Stefanski and Carroll [29,30], Liu and Taylor [22], Fan [12–14] and Ruymgaart [28]. Some examples for deconvolution problems are given in Carroll and Hall [4], a real practical application is described in Mendelsohn and Rice [26]. An amusing example of a problem with a perfectly known convolution operator is described in Fuller [16, p. 201].

E-mail address: mneumann@mathematik.uni-jena.de.

0047-259X/$ - see front matter
doi:10.1016/j.jmva.2006.09.012

On the other hand, apart from being quite convenient for mathematical tractability, the assumption that the distribution of the errors is perfectly known seems to be unrealistic in most practical applications. Sometimes one can draw some approximate information about it due to some additional source which allows to estimate it. In the case that information about the error distribution can be drawn from an additional experiment, Diggle and Hall [8] proposed to use the standard kernel deconvolution technique with the empirical characteristic function of the errors inserted for their unknown characteristic function. The effect of estimating the error density on rates of convergence and a modified regularization scheme has been studied by Neumann [27]; see also Efromovich [10] and Meister [23] for more work in this area. The case of a general unknown operator which can be estimated on the basis of training data has been considered by Efromovich and Koltchinskii [11] and Cavalier and Hengartner [5]. Some special cases where the error density and the distribution of interest have different characteristics and can therefore be both identified are considered by Butucea and Matias [3] and Meister [24,25]. Butucea and Matias [3] and Meister [25] considered the case with an ordinary smooth density and a supersmooth error density while Meister [24] investigated the case with an ordinary smooth density and two possible error densities, an ordinary smooth and a supersmooth one. In all of these cases the true error distribution can be identified from the tail behavior of the characteristic function of the observations.

Another important instance, where consistent deconvolution without prior knowledge of the error distribution and even without a training set for estimating the error distribution is possible, is the case of panel data. In this case, data from the distribution of interest are repeatedly observed, each time with an independent error. Horowitz and Markatou [18] studied the case with observations essentially of the type

$$Y_{ij} = X_i + \varepsilon_{ij}, \quad i = 1, \ldots, n, \quad j = 1, \ldots, N, \tag{1.1}$$

where all random variables involved are independent and $X_i \sim P_X$, $\varepsilon_{ij} \sim P_\varepsilon$. They proposed to estimate first the characteristic function φ_ε of P_ε and to divide then the empirical characteristic function of the Y_{ij}'s by that estimate and regularize this by a kernel function. Consistent estimation of φ_ε is obviously possible if $N \geqslant 2$ and the distribution P_ε is symmetric about 0 since then φ_ε is just the square root of the characteristic function of $Y_{i1} - Y_{i2}$. In the general case of a not necessarily symmetric error distribution, they outlined a possibly consistent method of estimating φ_ε in the case of $N \geqslant 3$, however, a rigorous asymptotic theory is lacking. Under the assumption that the characteristic functions φ_X and φ_ε do not vanish, Li and Vuong [21] proved that they can actually be identified up to a location shift from the characteristic function of $(Y_{i1}, \ldots, Y_{iN})'$ and proposed an estimator of the densities of X_i and ε_{ij}. Moreover, in the special case that the modulii of the characteristic functions have a regular (polynomial or exponential) decay, they gave rates for an appropriate choice of a spectral cut-off parameter and derived rates of convergence of the density estimators. Assuming that the cumulative distribution functions of P_X and P_ε are continuous and that moments of sufficiently high order are finite Hall and Yao [17] also proposed consistent estimators of the characteristic functions and the densities of the involved random variables.

In the present paper, we study again the case of panel data obeying (1.1). We assume that $N \geqslant 2$ and try to avoid as far as possible any assumption on the distribution of interest P_X and the error distribution P_ε. Since a rigorous asymptotic theory for explicit estimators of φ_ε seems to be rather cumbersome in our quite general setting we devise a completely different approach. It is clear that the characteristic function φ_Z of $Z_i = (Y_{i1}, \ldots, Y_{iN})'$ can be consistently estimated by its empirical version $\widehat{\varphi}_{Z,n}$. In view of the fact that $\varphi_Z(\omega_1, \ldots, \omega_N) = \varphi_X(\omega_1 + \cdots + \omega_N)\varphi_\varepsilon(\omega_1) \cdots \varphi_\varepsilon(\omega_N)$

holds for all $\omega_1, \ldots, \omega_N$, we will fit a pair of characteristic functions, $\widehat{\varphi}_{X,n}$ and $\widehat{\varphi}_{\varepsilon,n}$, to $\widehat{\varphi}_{Z,n}$ by a minimum distance method and take the corresponding distributions $\widehat{P}_{X,n}$ and $\widehat{P}_{\varepsilon,n}$ as estimators of P_X and P_ε, respectively. These estimators are actually consistent since we can show that φ_Z uniquely determines P_X and P_ε under side conditions such as $\mathrm{median}(P_\varepsilon) = 0$ or $E_{P_\varepsilon}\varepsilon = 0$. An interesting aspect of our approach is that we need not explicitly invert the convolution mapping. We think that such an approach is of potential interest also in other cases of ill-posed statistical inverse problems where an analytic inversion of the operator is difficult. Moreover, since we intend to estimate the distributions P_X and P_ε but not their densities (if they exist at all) our method does not involve any smoothing parameter.

2. Assumptions and main results

We assume that we observe Y_{ij} $(j = 1, \ldots, N;\, i = 1, \ldots, n)$ obeying the equation

$$Y_{ij} = X_i + \varepsilon_{ij}, \tag{2.1}$$

where all random variables appearing on the right-hand side of (2.1) are independent. We intend to estimate the common distribution P_X of the X_i's. We do not assume that any prior knowledge about the common distribution P_ε of the ε_{ij}'s is available, eventually apart from an obviously necessary identifiability condition. Denote by φ_X and φ_ε the characteristic functions of P_X and P_ε, respectively. It is clear that there were no chance of a consistent estimator in the case of no replications, $N = 1$. In the case of replications, however, consistent estimation of P_X and P_ε is possible under certain circumstances. Horowitz and Markatou [18] described a consistent method under the additional assumption that P_ε is symmetric around 0 and $N \geqslant 2$ and sketched an idea of a possibly consistent procedure in the general case if $N \geqslant 3$. Their method amounts to first estimating φ_ε and then plugging this estimator into a standard deconvoluting kernel estimator. A more thorough study of this problem was undertaken by Li and Vuong [21] who showed that φ_X and φ_ε are actually identifiable if these characteristic functions do not vanish and proved consistency for corresponding density estimators under the stronger condition of a regular decay of the modulii of the characteristic functions.

We will devise consistent estimators of P_X and P_ε in the case that at least two replications $(N \geqslant 2)$ are available and try to do this under minimal conditions on the null sets of the characteristic functions involved. As a starting point, we estimate the characteristic function φ_Z of $Z_i = (Y_{i1}, \ldots, Y_{iN})'$ by

$$\widehat{\varphi}_{Z,n}(\omega_1, \ldots, \omega_N) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{N!} \sum_{1 \leqslant j_1, \ldots, j_N \leqslant N:\, j_k \neq j_l \text{ for } k \neq l} e^{i(\omega_1 Y_{ij_1} + \cdots + \omega_N Y_{ij_N})}. \tag{2.2}$$

It follows from the multivariate Glivenko-Cantelli theorem that

$$P\left(\widehat{\varphi}_{Z,n}(\omega_1, \ldots, \omega_N) \underset{n \to \infty}{\to} \varphi_Z(\omega_1, \ldots, \omega_N) \quad \forall \omega_1, \ldots, \omega_N \right) = 1, \tag{2.3}$$

see also Theorem 3.2.1 in Ushakov [31, p. 165].

The basic reason why consistent estimation of P_X and P_ε will be possible is that φ_Z completely determines P_X and P_ε under certain circumstances. Before we formulate such a result, we have to exclude some cases where identifiability of P_X and P_ε cannot be guaranteed. First, since the observations $(Y_{i1}, \ldots, Y_{iN})'$, $i = 1, \ldots, N$, retain their common distribution if P_X and P_ε are replaced by the shifted distributions $P_X(\cdot - c)$ and $P_\varepsilon(\cdot + c)$, for any $c \in \mathbb{R}$, it is clear that we can at best identify them up to a location shift unless an additional condition regarding the location of P_ε (or of P_X) is stipulated. For reasons of clarity of presentation, we consider first the case without such a condition and seek conditions under which P_X and P_ε can be identified from φ_Z up to a location parameter.

As an instructive example, we consider the distribution $\bar{P}_\varepsilon$ with characteristic function $\bar{\varphi}_\varepsilon(\omega) = (1-|\omega|)_+$. Now we can choose two different distributions $\bar{P}_{X,1}$ and $\bar{P}_{X,2}$ whose respective characteristic functions satisfy $\bar{\varphi}_{X,1}(\omega) = \bar{\varphi}_{X,2}(\omega)\ \forall|\omega| \leqslant N$; see for example Feller [15, p. 479]. Then it follows that $\bar{\varphi}_{X,1}(\omega_1+\cdots+\omega_N)\bar{\varphi}_\varepsilon(\omega_1)\cdots\bar{\varphi}_\varepsilon(\omega_N) = \bar{\varphi}_{X,2}(\omega_1+\cdots+\omega_N)\bar{\varphi}_\varepsilon(\omega_1)\cdots\bar{\varphi}_\varepsilon(\omega_N)$ $\forall\omega_1,\ldots,\omega_N \in \mathbb{R}$, that is, we cannot consistently distinguish between the cases of $X_i \sim \bar{P}_{X,1}$, $\varepsilon_{ij} \sim \bar{P}_\varepsilon$ and $X_i \sim \bar{P}_{X,2}$, $\varepsilon_{ij} \sim \bar{P}_\varepsilon$. We learn from this example that we have to exclude cases where φ_ε vanishes on a too large domain. Since any closed set A with $0 \notin A$ and $-A = A$ can be the zero set of a characteristic function (see [19, Corollary 1]) we will impose exactly those properties of the zero sets of φ_ε and φ_X that we use in the proof of Lemma 2.1.

(A1) (i) If $N = 2$, then the set $\{\omega : \varphi_\varepsilon(\omega 2^{-k}) \neq 0$ and $\varphi_X(\omega 2^{-k}) \neq 0 \quad \forall k = 0,1,\ldots\}$ is assumed to be dense in $\mathbb{R}$.
(ii) If $N \geqslant 3$, then the set $\{\omega : \varphi_\varepsilon(\omega(N-1)^{-k}) \neq 0 \quad \forall k = 0,1,\ldots\}$ is assumed to be dense in $\mathbb{R}$.

We think that this assumption is not very restrictive. It excludes characteristic functions that vanish on nonempty open subsets of $\mathbb{R}$. Most textbook distributions, however, have a characteristic function with at most countably many zeros, hence, satisfying also (A1).

The following lemma states that knowledge of φ_Z actually suffices to identify P_X and P_ε, up to a location shift. It extends Lemma 1 in Kotlarski [20] in that we do not require that the involved characteristic functions do not vanish.

Lemma 2.1. *Suppose that P_X and P_ε are distributions with characteristic functions φ_X and φ_ε satisfying* (A1). *Let $\widetilde{P}_X$ and $\widetilde{P}_\varepsilon$ be further distributions with respective characteristic functions $\widetilde{\varphi}_X$ and $\widetilde{\varphi}_\varepsilon$. If now*

$$\begin{aligned}&\varphi_X(\omega_1+\cdots+\omega_N)\varphi_\varepsilon(\omega_1)\cdots\varphi_\varepsilon(\omega_N)\\&\quad=\widetilde{\varphi}_X(\omega_1+\cdots+\omega_N)\widetilde{\varphi}_\varepsilon(\omega_1)\cdots\widetilde{\varphi}_\varepsilon(\omega_N)\quad\forall\omega_1,\ldots,\omega_N\in\mathbb{R},\end{aligned}\tag{2.4}$$

then there exists a constant $c \in \mathbb{R}$ such that

$$\widetilde{P}_X = P_{X+c} \quad \text{and} \quad \widetilde{P}_\varepsilon = P_{\varepsilon-c},$$

that is, P_X and $\widetilde{P}_X$ as well as P_ε and $\widetilde{P}_\varepsilon$ are equal up to a location shift.

Estimators $\widehat{P}_{X,n}$ and $\widehat{P}_{\varepsilon,n}$ of P_X and P_ε will be defined via a minimum distance fit in the frequency domain. Recall that $\varphi_Z(\omega_1,\ldots,\omega_N) = \varphi_X(\omega_1+\cdots+\omega_N)\varphi_\varepsilon(\omega_1)\cdots\varphi_\varepsilon(\omega_N)$. Let $K : \mathbb{R}^N \to (0,\infty)$ be any continuous and everywhere positive probability density. We define a distance ρ as

$$\begin{aligned}&\rho(\widetilde{\varphi}_X,\widetilde{\varphi}_\varepsilon;\widetilde{\varphi}_Z)\\&\quad=\int_{\mathbb{R}^N}\left|\widetilde{\varphi}_X(\omega_1+\cdots+\omega_N)\widetilde{\varphi}_\varepsilon(\omega_1)\cdots\widetilde{\varphi}_\varepsilon(\omega_N)-\widetilde{\varphi}_Z(\omega_1,\ldots,\omega_N)\right|K(\omega_1,\ldots,\omega_N)\,d\omega.\end{aligned}\tag{2.5}$$

Now we intend to define $\widehat{P}_{X,n}$ and $\widehat{P}_{\varepsilon,n}$ such that their characteristic functions $\widehat{\varphi}_{X,n}$ and $\widehat{\varphi}_{\varepsilon,n}$ minimize $\rho(\cdot,\cdot;\widehat{\varphi}_{Z,n})$. Since we cannot guarantee that the infimum will actually be attained we choose first a vanishing sequence $(\delta_n)_{n\in\mathbb{N}}$ and then probability distributions $\widehat{P}_{X,n}$ and $\widehat{P}_{\varepsilon,n}$ such that the corresponding characteristic functions $\widehat{\varphi}_{X,n}$ and $\widehat{\varphi}_{\varepsilon,n}$ fulfill

$$\rho\left(\widehat{\varphi}_{X,n}, \widehat{\varphi}_{\varepsilon,n}; \widehat{\varphi}_{Z,n}\right) \leqslant \inf_{\widetilde{\varphi}_X, \widetilde{\varphi}_\varepsilon \in \Phi} \rho\left(\widetilde{\varphi}_X, \widetilde{\varphi}_\varepsilon; \widehat{\varphi}_{Z,n}\right) + \delta_n, \tag{2.6}$$

where $\Phi = \{\varphi : \varphi$ is a characteristic function of a probability distribution$\}$. Note that we do not require that $\widehat{\varphi}_{X,n}$ and $\widehat{\varphi}_{\varepsilon,n}$ satisfy (A1). This approach bears similarities with a method proposed by Hall and Yao [17]. They also considered a minimum distance fit and used histograms to generate estimators of P_X and P_ε. The method investigated here is slightly different and is shown to be consistent under weaker conditions on the involved distributions.

Denote by $\widehat{P}_{(X+\varepsilon_1,\ldots,X+\varepsilon_N)}$ the distribution with characteristic function $\widehat{\varphi}_{X,n}(\omega_1 + \cdots + \omega_n)$ $\widehat{\varphi}_{\varepsilon,n}(\omega_1)\cdots\widehat{\varphi}_{\varepsilon,n}(\omega_N)$. Before we state consistency properties for the sequences $(\widehat{P}_{X,n})_{n\in\mathbb{N}}$ and $(\widehat{P}_{\varepsilon,n})_{n\in\mathbb{N}}$, we introduce a metric which metrizes weak convergence. For probability measures P_n with respective cumulative distribution functions F_n, define the Lévy distance as

$$d(P_1, P_2) = \inf\{\varepsilon : F_1(x-\varepsilon) - \varepsilon \leqslant F_2(x) \leqslant F_1(x+\varepsilon) + \varepsilon \quad \forall x \in \mathbb{R}\}.$$

It is known that $P_n \Longrightarrow P_0$ if and only if $d(F_n, F_0) \underset{n\to\infty}{\longrightarrow} 0$; see for example Chung [6, p. 94]. Now we define, for probability measures $\widetilde{P}_X$ and $\widetilde{P}_\varepsilon$,

$$\Delta(\widetilde{P}_X, \widetilde{P}_\varepsilon; P_X, P_\varepsilon) = \inf_{c\in\mathbb{R}} \left\{d(\widetilde{P}_X, P_{X+c}) + d(\widetilde{P}_\varepsilon, P_{\varepsilon-c})\right\}.$$

The following theorem states some sort of consistency property of the estimator sequences, of course, up to a location shift.

Theorem 2.1. *Suppose that observations according to* (2.1) *are given. Then, as $n \to \infty$,*

(i) $\widehat{P}_{(X+\varepsilon_1,\ldots,X+\varepsilon_N)} \Longrightarrow P_Z$ *almost surely.*
(ii) *If additionally assumption* (A1) *is fulfilled, then*

$$\Delta\left(\widehat{P}_{X,n}, \widehat{P}_{\varepsilon,n}; P_X, P_\varepsilon\right) \xrightarrow{\text{a.s}} 0.$$

As already indicated, to make the distributions P_X and P_ε identifiable we have to impose some additional condition. Since the ε_{ij}'s are interpreted as errors it is natural to assume that they are centered in some sense. In what follows we will stipulate the following conditions:

(A2) P_ε obeys one of the following conditions:
  (i) median$(P_\varepsilon) = 0$ and $P_\varepsilon((-\infty, -\delta]) < \frac{1}{2} < P_\varepsilon((-\infty, \delta))\ \forall \delta > 0$,
  (ii) $E\varepsilon_{ij} = 0$.

If (A2)(i) is stipulated, then we obtain consistent estimators of P_X and P_ε by a minimum distance fit under a corresponding side condition on the median. Denote $\Phi^{(i)} = \{\varphi : \varphi$ is a characteristic function of a probability distribution P with median 0$\}$. Note that $\Phi^{(i)}$ includes distributions whose median is not uniquely defined; uniqueness of the median is actually only required for the true error distribution P_ε, under (A2)(i). Now we choose $\widehat{\widehat{P}}_{X,n}$ and $\widehat{\widehat{P}}_{\varepsilon,n}$ such that their characteristic functions fulfill $\widehat{\widehat{\varphi}}_{X,n} \in \Phi$, $\widehat{\widehat{\varphi}}_{\varepsilon,n} \in \Phi^{(i)}$ and

$$\rho\left(\widehat{\widehat{\varphi}}_{X,n}, \widehat{\widehat{\varphi}}_{\varepsilon,n}; \widehat{\varphi}_{Z,n}\right) \leqslant \inf_{\widetilde{\varphi}_X \in \Phi, \widetilde{\varphi}_\varepsilon \in \Phi^{(i)}} \rho\left(\widetilde{\varphi}_X, \widetilde{\varphi}_\varepsilon; \widehat{\varphi}_{Z,n}\right) + \delta_n. \tag{2.7}$$

Again, we do not require that the characteristic functions $\widehat{\widehat{\varphi}}_{X,n}$, $\widehat{\widehat{\varphi}}_{\varepsilon,n}$ satisfy (A1).

In the second case, we have to be more careful. The side condition $E_{\widehat{\widehat{P}}_{\varepsilon,n}}\varepsilon = 0$ does not suffice to guarantee that any eventually existing weak limit of $(\widehat{\widehat{P}}_{\varepsilon,n})_{n\in\mathbb{N}}$ has also expectation 0. This, however, would be the case if we could additionally ensure uniform integrability of the random variables $\widehat{\widehat{\varepsilon}}_n \sim \widehat{\widehat{P}}_{\varepsilon,n}$. This, in turn, would follow if we had, besides weak convergence, that $E_{\widehat{\widehat{P}}_{\varepsilon,n}}|\varepsilon| \xrightarrow{\text{a.s}} E_{P_\varepsilon}|\varepsilon|$. Since $E_{P_\varepsilon}|\varepsilon|$ is not identifiable from the observations in our setting we enforce instead of that uniform integrability of the random variables $\widehat{\widehat{\varepsilon}}_{n,1} - \widehat{\widehat{\varepsilon}}_{n,2}$, where $\widehat{\widehat{\varepsilon}}_{n,1}$ and $\widehat{\widehat{\varepsilon}}_{n,2}$ are independent with distribution $\widehat{\widehat{P}}_{\varepsilon,n}$. It will be shown that this implies uniform integrability of $(\widehat{\widehat{\varepsilon}}_{n,1})_{n\in\mathbb{N}}$ and, therefore, the desired centering of the limit distribution.

Denote $\mu_1 = E|\varepsilon_{i1} - \varepsilon_{i2}|$. A consistent estimator of μ_1 is given by

$$\widehat{\mu}_{1,n} = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{\binom{N}{2}} \sum_{1\leqslant j_1<j_2\leqslant N} |Y_{ij_1} - Y_{ij_2}|.$$

Let $\Phi_n^{(ii)} = \{\varphi : \varphi$ is a characteristic function of a probability measure P, $E_P\varepsilon_1 = 0$ and $E_P|\varepsilon_1 - \varepsilon_2| \leqslant \widehat{\mu}_{1,n}$, for independent $\varepsilon_1, \varepsilon_2 \sim P\}$. Now we choose $\widehat{\widehat{P}}_{X,n}$ and $\widehat{\widehat{P}}_{\varepsilon,n}$ such that their characteristic functions fulfill $\widehat{\widehat{\varphi}}_{X,n} \in \Phi$, $\widehat{\widehat{\varphi}}_{\varepsilon,n} \in \Phi_n^{(ii)}$ and

$$\rho\left(\widehat{\widehat{\varphi}}_{X,n}, \widehat{\widehat{\varphi}}_{\varepsilon,n}; \widehat{\varphi}_{Z,n}\right) \leqslant \inf_{\widetilde{\varphi}_X\in\Phi, \widetilde{\varphi}_\varepsilon\in\Phi_n^{(ii)}} \rho\left(\widetilde{\varphi}_X, \widetilde{\varphi}_\varepsilon; \widehat{\varphi}_{Z,n}\right) + \delta_n. \tag{2.8}$$

Loosely speaking, the side condition $E_{\widehat{\widehat{P}}_{\varepsilon,n}}|\varepsilon_1 - \varepsilon_2| \leqslant \widehat{\mu}_{1,n}$ is used to "squeeze out" any asymptotically superfluous part of the distributions $(\widehat{\widehat{P}}_{\varepsilon,n})_{n\in\mathbb{N}}$. If we can show that $\widehat{\widehat{\varepsilon}}_{n,1} - \widehat{\widehat{\varepsilon}}_{n,2} \xrightarrow{d} \varepsilon_{11} - \varepsilon_{12}$, then we can conclude that $E_{\widehat{\widehat{P}}_{\varepsilon,n}}|\widehat{\widehat{\varepsilon}}_{n,1} - \widehat{\widehat{\varepsilon}}_{n,2}| \underset{n\to\infty}{\longrightarrow} E_{P_\varepsilon}|\varepsilon_{11} - \varepsilon_{12}|$, which implies uniform integrability of $(\widehat{\widehat{\varepsilon}}_{n,1} - \widehat{\widehat{\varepsilon}}_{n,2})_{n\in\mathbb{N}}$. The following theorem states that we obtain strong consistency under either one of the identifiability conditions in (A2).

Theorem 2.2. *Suppose that observations according to* (2.1) *are given and that assumptions* (A1) *and* (A2) *are fulfilled. Then, as* $n \to \infty$,

$$d(\widehat{\widehat{P}}_{X,n}, P_X) \xrightarrow{\text{a.s}} 0 \quad \textit{and} \quad d(\widehat{\widehat{P}}_{\varepsilon,n}, P_\varepsilon) \xrightarrow{\text{a.s}} 0.$$

Remark 1. (i) *Practical computation*: When implementing the procedure we face the problem that there is no way of computing $\widehat{P}_{X,n}$ directly. For example, we cannot define $\widehat{\varphi}_{X,n}$ and $\widehat{\varphi}_{\varepsilon,n}$ as pointwise minimizer of $|\widehat{\varphi}_{X,n}(\omega_1 + \cdots + \omega_N)\widehat{\varphi}_{\varepsilon,n}(\omega_1)\cdots\widehat{\varphi}_{\varepsilon,n}(\omega_N) - \widehat{\varphi}_{Z,n}(\omega_1, \ldots, \omega_N)|$ since we have to regard the side condition that $\widehat{\varphi}_{X,n}$ and $\widehat{\varphi}_{\varepsilon,n}$ must be characteristic functions. To replace the infinite-dimensional minimization problem (2.6) (as well as (2.7) and (2.8)) by a finite-dimensional one, we can make use of the fact that any distribution can be represented as the weak limit of discrete distributions with a finite number of atoms. (For example, by Glivenko-Cantelli, the empirical distributions converge to the underlying distribution as the sample size tends to infinity.) So we may choose $\widehat{P}_{X,n}$ and $\widehat{P}_{\varepsilon,n}$ from a family of distributions $\{P = \sum_{i=1}^{M} \alpha_i\delta_{x_i} : \alpha_i \geqslant 0, \sum_{i=1}^{M}\alpha_i = 1, x_i \in \mathbb{R}\}$, where δ_x denotes the Dirac measure at x. Consistency is then possible if $M = M(n) \to \infty$, as $n \to \infty$. We think that a feasible algorithm can be derived along the lines of the simulating annealing approach which was employed by Hall and Yao [17] in a similar context. If it is conjectured that the cumulative distribution functions of P_X and P_ε are smooth, it makes also sense to use a sieve of smooth distributions.

(ii) *Rates of convergence*: An interesting approach different to ours has been investigated by Li and Vuong [21]. Under the assumption of nowhere vanishing characteristic functions φ_X and φ_ε, they first showed that these functions can be obtained from the characteristic function ψ of $(Y_{i1}, Y_{i2})'$ as

$$\varphi_X(\omega) = \exp \int_0^\omega \frac{\partial\psi(0, u_2)/\partial u_1}{\psi(0, u_2)} du_2,$$
$$\varphi_\varepsilon(\omega) = \frac{\psi(0, \omega)}{\varphi_X(\omega)}.$$

Starting with the empirical characteristic function of $(Y_{i1}, Y_{i2})'$ they used these formulas to construct estimates of φ_X and φ_ε and finally obtained estimates of the densities of the X_i and the ε_{ij} by a windowed inverse Fourier transform. In the special case that the modulii of φ_X and φ_ε have a regular (that is, polynomial or exponential) decay, they gave rates for an appropriate choice of the spectral cut-off parameter and derived rates of convergence for their density estimators.

Our intention here is to devise consistent estimators of P_X and P_ε under minimal conditions. The assumption of a regular decay of the modulus of the characteristic function is met only by a few textbook distributions; the normal, Cauchy, gamma and double exponential distributions are usually mentioned examples. Moreover, even Li and Vuong's [21] basic assumption of nowhere vanishing characteristic functions is not fulfilled by several textbook distributions. Examples can be found in Ushakov [31, Appendix B]; the arcsine, discrete and continuous uniform and triangular distributions are among them. Without explicit expressions characterizing the dependency of φ_X and φ_ε on φ_Z as above, however, it seems to be rather difficult to characterize the degree of ill-posedness of the deconvolution problem which would be a first step in deriving sharp rates of convergence. In view of this, we think that reasonable results for rates of convergence are beyond reach in our general context.

(iii) *Limitations*: If the cumulative distribution functions of P_X and P_ε are continuous, then it follows from the weak convergence results in Theorem 2.2 that

$$\sup_{x\in\mathbb{R}} \left| \widehat{\widehat{P}}_{X,n}((-\infty, x]) - P_X((-\infty, x]) \right| \xrightarrow{\text{a.s}} 0$$

and

$$\sup_{x\in\mathbb{R}} \left| \widehat{\widehat{P}}_{\varepsilon,n}((-\infty, x]) - P_\varepsilon((-\infty, x]) \right| \xrightarrow{\text{a.s}} 0.$$

This is, however, in general not true if one of the cumulative distribution functions is not continuous. Consider, for example, the case where $N = 2$, $P_\varepsilon = \mathcal{N}(0, 1)$ and $P_X \in \{P_{X,1}^{(n)}, P_{X,2}^{(n)}\}$, $P_{X,1}^{(n)} = (\delta_1 + \delta_{-1})/2$, $P_{X,2}^{(n)} = (\delta_{1+c_n} + \delta_{-1-c_n})/2$. The random vector $Z = (X_1 + \varepsilon_{11}, X_1 + \varepsilon_{12})'$ has respective distributions $P_{Z,1}^{(n)}$ and $P_{Z,2}^{(n)}$ which are both mixtures of four bivariate normal distributions. Moreover, it follows for the Hellinger distance that

$$H^2(P_{Z,1}^{(n)}, P_{Z,2}^{(n)}) = O(n^{-1}),$$

if $c_n = O(n^{-1/2})$. Hence, it follows from standard arguments (see, for example, [9]) that one cannot consistently distinguish between the two cases $X_i \sim P_{X,1}^{(n)}$ and $X_i \sim P_{X,2}^{(n)}$, as $n \to \infty$, which in turn means that there does not exist any uniformly consistent (w.r.t the Kolmogorov distance) estimator of P_X.

3. Proofs

Proof of Lemma 2.1. The first part of the proof is analogous to the proof of Lemma 1 in Kotlarski [20]; we include it for the sake of self-consistency of our proof.

Since φ_X and φ_ε are characteristic functions there exists an $\omega_0 > 0$ such that $\varphi_\varepsilon(\omega) \neq 0$ and $\varphi_X(2\omega) \neq 0$ if $|\omega| \leqslant \omega_0$. Define, for $|\omega| \leqslant 2\omega_0$,

$$p_X(\omega) = \widetilde{\varphi}_X(\omega)/\varphi_X(\omega).$$

p_X is a continuous complex function which is equal to 1 at 0.

It follows from (2.4), for $\omega_1, \omega_2 \in [-\omega_0, \omega_0]$, that

$$\begin{aligned}\frac{\varphi_X(\omega_1+\omega_2)}{\varphi_X(\omega_1)\varphi_X(\omega_2)} &= \frac{\varphi_X(\omega_1+\omega_2)\varphi_\varepsilon(\omega_1)\varphi_\varepsilon(\omega_2)}{\varphi_X(\omega_1+0)\varphi_\varepsilon(\omega_1)\varphi_\varepsilon(0)\ \varphi_X(\omega_2+0)\varphi_\varepsilon(\omega_2)\varphi_\varepsilon(0)} \\ &= \frac{\widetilde{\varphi}_X(\omega_1+\omega_2)\widetilde{\varphi}_\varepsilon(\omega_1)\widetilde{\varphi}_\varepsilon(\omega_2)}{\widetilde{\varphi}_X(\omega_1+0)\widetilde{\varphi}_\varepsilon(\omega_1)\widetilde{\varphi}_\varepsilon(0)\ \widetilde{\varphi}_X(\omega_2+0)\widetilde{\varphi}_\varepsilon(\omega_2)\widetilde{\varphi}_\varepsilon(0)} \\ &= \frac{\widetilde{\varphi}_X(\omega_1+\omega_2)}{\widetilde{\varphi}_X(\omega_1)\widetilde{\varphi}_X(\omega_2)},\end{aligned} \tag{3.1}$$

which implies

$$p_X(\omega_1+\omega_2) = p_X(\omega_1)p_X(\omega_2) \quad \forall \omega_1, \omega_2 \in [-\omega_0, \omega_0]. \tag{3.2}$$

This is a so-called Cauchy equation. We obtain from Theorem 2.1.4.1 in Aczél [1, p. 46] that the only continuous solution of (3.2) satisfying the side condition $p_X(0) = 1$ is given by

$$p_X(\omega) = e^{b\omega},$$

where b is any complex number. Since $p_X(-\omega) = \overline{p_X(\omega)}$ we see that $b = ic$, for some real c. Therefore, we conclude that

$$\widetilde{\varphi}_X(\omega) = e^{ic\omega}\varphi_X(\omega) \quad \forall \omega \in [-2\omega_0, 2\omega_0]. \tag{3.3}$$

Furthermore, (2.4) yields immediately that

$$\widetilde{\varphi}_\varepsilon(\omega) = e^{-ic\omega}\varphi_\varepsilon(\omega) \quad \forall \omega \in [-2\omega_0, 2\omega_0]. \tag{3.4}$$

Now it remains to extend Eqs. (3.3) and (3.4) to the whole real line. From here on we have to distinguish between the two cases $N = 2$ and $N \geqslant 3$.

(i) $N = 2$: Let $\omega \in A := \{\omega : \varphi_\varepsilon(\omega 2^{-k}) \neq 0 \text{ and } \varphi_X(\omega 2^{-k}) \neq 0 \quad \forall k = 0, 1, \ldots\}$ be arbitrary. We obtain, analogously to (3.1), that

$$\frac{\varphi_X(\omega)}{(\varphi_X(\omega/2))^2} = \frac{\widetilde{\varphi}_X(\omega)}{(\widetilde{\varphi}_X(\omega/2))^2}.$$

Iterating this scheme we get

$$\frac{\varphi_X(\omega)}{(\varphi_X(\omega 2^{-K}))^{2^K}} = \frac{\widetilde{\varphi}_X(\omega)}{(\widetilde{\varphi}_X(\omega 2^{-K}))^{2^K}}.$$

Using this equation with a K large enough such that $|\omega 2^{-K}| \leqslant 2\omega_0$ we conclude from (3.3) that

$$\widetilde{\varphi}_X(\omega) = \varphi_X(\omega)\left(\frac{\widetilde{\varphi}_X(\omega 2^{-K})}{\varphi_X(\omega 2^{-K})}\right)^{2^K} = \varphi_X(\omega)\, e^{ic\omega}.$$

Since A is dense in $\mathbb{R}$ and $\{\omega : \widetilde{\varphi}_X(\omega) = e^{ic\omega}\varphi_X(\omega)\}$ is a closed set we conclude that

$$\widetilde{\varphi}_X(\omega) = e^{ic\omega}\varphi_X(\omega) \quad \forall \omega \in \mathbb{R},$$

that is, $\widetilde{P}_X = P_{X+c}$. This implies, again by (2.4), that

$$\widetilde{\varphi}_\varepsilon(\omega) = e^{-ic\omega}\varphi_\varepsilon(\omega) \quad \forall \omega \in \mathbb{R},$$

which yields $\widetilde{P}_\varepsilon = P_{\varepsilon-c}$.

(ii) $N \geqslant 3$: Let now $\omega \in B := \{\omega : \varphi_\varepsilon(\omega(N-1)^{-k}) \neq 0\, \forall k = 0, 1, \ldots\}$ be arbitrary. Choosing in (2.4) $\omega_1 = \omega$ and $\omega_2 = \cdots = \omega_N = -\omega/(N-1)$ we obtain

$$\varphi_\varepsilon(\omega)\left(\varphi_\varepsilon(-\omega/(N-1))\right)^{N-1} = \widetilde{\varphi}_\varepsilon(\omega)\left(\widetilde{\varphi}_\varepsilon(-\omega/(N-1))\right)^{N-1}.$$

Iterating this scheme we obtain that

$$\frac{\varphi_\varepsilon(\omega)}{\left(\varphi_\varepsilon(\omega/(-(N-1))^K)\right)^{(-(N-1))^K}} = \frac{\widetilde{\varphi}_\varepsilon(\omega)}{\left(\widetilde{\varphi}_\varepsilon(\omega/(-(N-1))^K)\right)^{(-(N-1))^K}}.$$

Using this equation with a K large enough such that $|\omega/(N-1)^K| \leqslant 2\omega_0$ we obtain from (3.4) that

$$\widetilde{\varphi}_\varepsilon(\omega) = \varphi_\varepsilon(\omega)\left(\frac{\widetilde{\varphi}_\varepsilon(\omega/(-(N-1))^K)}{\varphi_\varepsilon(\omega/(-(N-1))^K)}\right)^{(-(N-1))^K} = \varphi_\varepsilon(\omega)\, e^{-ic\omega}.$$

Since B is dense in $\mathbb{R}$ and $\{\omega : \widetilde{\varphi}_\varepsilon(\omega) = e^{-ic\omega}\varphi_\varepsilon(\omega)\}$ is a closed set we conclude that

$$\widetilde{\varphi}_\varepsilon(\omega) = e^{-ic\omega}\varphi_\varepsilon(\omega) \quad \forall \omega \in \mathbb{R},$$

that is, $\widetilde{P}_\varepsilon = P_{\varepsilon-c}$. Using once more (2.4) this implies

$$\widetilde{\varphi}_X(\omega) = e^{ic\omega}\,\varphi_X(\omega) \quad \forall \omega \in \mathbb{R},$$

hence, $\widetilde{P}_X = P_{X+c}$. $\square$

Proof of Theorem 2.1. (i) From the pointwise convergence of $\widehat{\varphi}_{Z,n}$ to φ_Z stated in (2.3), we obtain by the dominated convergence theorem that

$$\rho(\varphi_X, \varphi_\varepsilon; \widehat{\varphi}_{Z,n}) \stackrel{\text{a.s}}{\longrightarrow} 0. \tag{3.5}$$

By (2.6) we have

$$\rho\left(\widehat{\varphi}_{X,n}, \widehat{\varphi}_{\varepsilon,n}; \widehat{\varphi}_{Z,n}\right) \leqslant \rho\left(\varphi_X, \varphi_\varepsilon; \widehat{\varphi}_{Z,n}\right) + \delta_n,$$

which implies that

$$\begin{aligned}\rho\left(\widehat{\varphi}_{X,n}, \widehat{\varphi}_{\varepsilon,n}; \varphi_Z\right) &\leqslant \rho\left(\widehat{\varphi}_{X,n}, \widehat{\varphi}_{\varepsilon,n}; \widehat{\varphi}_{Z,n}\right) + \rho\left(\varphi_X, \varphi_\varepsilon; \widehat{\varphi}_{Z,n}\right) \\ &\leqslant 2\rho\left(\varphi_X, \varphi_\varepsilon; \widehat{\varphi}_{Z,n}\right) + \delta_n \quad \stackrel{\text{a.s}}{\longrightarrow} \quad 0.\end{aligned} \tag{3.6}$$

The rest of the proof follows a typical pattern in the context of weak convergence. We fix an arbitrary elementary event of the underlying probability space such that the convergence in (3.6) takes place. Then it follows that

$$\int_{s_1}^{t_1} \cdots \int_{s_N}^{t_N} \widehat{\varphi}_{X,n}(\omega_1 + \cdots + \omega_N)\widehat{\varphi}_{\varepsilon,n}(\omega_1) \cdots \widehat{\varphi}_{\varepsilon,n}(\omega_N)\, d\omega \underset{k\to\infty}{\longrightarrow} \int_{s_1}^{t_1} \cdots \int_{s_N}^{t_N} \varphi_X(\omega_1 + \cdots + \omega_N)\varphi_\varepsilon(\omega_1) \cdots \varphi_\varepsilon(\omega_N)\, d\omega \tag{3.7}$$

holds for all $-\infty < s_i < t_i < \infty$. On the left-hand side of (3.7), we have integrated characteristic functions of probability measures $\widehat{P}_{(X+\varepsilon_1,\ldots,X+\varepsilon_N),n}$ while we have on the right-hand side the integrated characteristic function of the probability measure P_Z. Since vague convergence implies pointwise convergence of the corresponding integrated characteristic functions and since these integrated characteristic functions are measure determining we can conclude that any sequence of subprobability distributions converges vaguely to some subprobability distribution if and only if the corresponding integrated characteristic functions converge pointwise; see also Theorem 6.3.3 in Chung [6] for such a result in the univariate case. Therefore, we obtain from (3.7) that

$$\widehat{P}_{(X+\varepsilon_1,\ldots,X+\varepsilon_N),n} \longrightarrow_{\mathrm{v}} P_Z.$$

Here '$\longrightarrow_{\mathrm{v}}$' denotes vague convergence to a possibly defective (that is, with a mass less than 1) measure on $\mathcal{B}$. Since this vague limit is a probability measure it follows that the mode of convergence is actually that of weak convergence, that is,

$$\widehat{P}_{(X+\varepsilon_1,\ldots,X+\varepsilon_N),n} \implies P_Z. \tag{3.8}$$

(ii) Now assume that additionally (A1) is fulfilled. It remains to prove that (3.8) implies

$$\Delta\left(\widehat{P}_{X,n}, \widehat{P}_{\varepsilon,n}; P_X, P_\varepsilon\right) \underset{n\to\infty}{\longrightarrow} 0. \tag{3.9}$$

To see this, assume that there exists a constant $\delta > 0$ and a subsequence $(n_k)_{k\in\mathbb{N}}$ of $\mathbb{N}$ such that

$$\Delta\left(\widehat{P}_{X,n_k}, \widehat{P}_{\varepsilon,n_k}; P_X, P_\varepsilon\right) \geqslant \delta \quad \forall k \in \mathbb{N}. \tag{3.10}$$

By Helly's selection theorem (see for example [2, p. 227]), there exist a further subsequence $(n'_k)_{k\in\mathbb{N}}$ of $(n_k)_{k\in\mathbb{N}}$ and subprobability distributions $P_{X,\infty}$ and $P_{\varepsilon,\infty}$ such that, as $k \to \infty$,

$$\widehat{P}_{X,n'_k} \longrightarrow_{\mathrm{v}} P_{X,\infty} \tag{3.11}$$

and

$$\widehat{P}_{\varepsilon,n'_k} \longrightarrow_{\mathrm{v}} P_{\varepsilon,\infty}. \tag{3.12}$$

Moreover, it follows from (3.8) that $(\widehat{P}_{X,n} * \widehat{P}_{\varepsilon,n})_{n\in\mathbb{N}}$ is a tight sequence of probability measures which implies that the measures $P_{X,\infty}$ and $P_{\varepsilon,\infty}$ have mass 1. Therefore, the modes of convergence in (3.11) and (3.12) are those of weak convergence, that is,

$$\widehat{P}_{X,n'_k} \implies P_{X,\infty}, \tag{3.13}$$

$$\widehat{P}_{\varepsilon,n'_k} \implies P_{\varepsilon,\infty} \tag{3.14}$$

and $P_{X,\infty}$ and $P_{\varepsilon,\infty}$ are probability measures on $\mathcal{B}$. Denote by $\varphi_{X,\infty}$ and $\varphi_{\varepsilon,\infty}$ the characteristic functions of $P_{X,\infty}$ and $P_{\varepsilon,\infty}$, respectively. Since weak convergence implies pointwise convergence of the corresponding characteristic functions we obtain by Fatou's lemma that

$$\rho\left(\varphi_{X,\infty}, \varphi_{\varepsilon,\infty}; \varphi_Z\right) \leqslant \liminf_{k\to\infty} \rho\left(\widehat{\varphi}_{X,n'_k}, \widehat{\varphi}_{\varepsilon,n'_k}; \varphi_Z\right) = 0.$$

Now we obtain by Lemma 2.1 that

$$P_{X,\infty} = P_{X+c} \quad \text{and} \quad P_{\varepsilon,\infty} = P_{\varepsilon-c}$$

hold for some $c \in \mathbb{R}$, which means in conjunction with (3.13) and (3.14) that

$$\Delta\left(\widehat{P}_{X,n'_k}, \widehat{P}_{\varepsilon,n'_k}; P_X, P_\varepsilon\right) \underset{k\to\infty}{\longrightarrow} 0.$$

Hence, our hypothesis (3.10) must be wrong, which completes the proof of the theorem. $\square$

Before we turn to the proof of Theorem 2.2, we prove an auxiliary result which implies that uniform integrability of $\widehat{\widehat{\varepsilon}}_{n,1}$ follows from uniform integrability of $\widehat{\widehat{\varepsilon}}_{n,1} - \widehat{\widehat{\varepsilon}}_{n,2}$.

Lemma 3.1. *Let ε_1 and ε_2 be independent and identically distributed random variables with mean zero. Then, for any $c > 0$,*

$$E\left[|\varepsilon_1 - \varepsilon_2| I\left(|\varepsilon_1 - \varepsilon_2| > \frac{c}{2}\right)\right] \geqslant \frac{1}{3} E\left[|\varepsilon_1| I\left(|\varepsilon_1| > c\right)\right].$$

Proof. Without loss of generality, we will show that

$$E\left[(\varepsilon_1 - \varepsilon_2)^+ I\left((\varepsilon_1 - \varepsilon_2)^+ > \frac{c}{2}\right)\right] \geqslant \frac{1}{3} E\left[\varepsilon_1^+ I\left(\varepsilon_1^+ > c\right)\right]. \tag{3.15}$$

(The corresponding inequality with $(\varepsilon_1 - \varepsilon_2)^-$ and ε_1^- follows analogously.) To prove (3.15), we distinguish between two cases.

If $P(\varepsilon_1 > c/2) < \frac{1}{3}$, then

$$\begin{aligned} E\left[(\varepsilon_1 - \varepsilon_2)^+ I\left((\varepsilon_1 - \varepsilon_2)^+ > \frac{c}{2}\right)\right] &\geqslant E\left[(\varepsilon_1 - \varepsilon_2)^+ I\left(\varepsilon_1 > c, \varepsilon_2 \leqslant \frac{c}{2}\right)\right] \\ &\geqslant E\left[\frac{\varepsilon_1}{2} I(\varepsilon_1 > c)\right] \cdot P\left(\varepsilon_2 \leqslant \frac{c}{2}\right) \\ &\geqslant \frac{1}{3}\, E\left[\varepsilon_1^+ I\left(\varepsilon_1^+ > c\right)\right]. \end{aligned}$$

If $P(\varepsilon_1 > c/2) \geqslant \frac{1}{3}$, then we obtain by $E[\varepsilon_2^-] = E[\varepsilon_1^-] = E[\varepsilon_1^+] \geqslant E[\varepsilon_1^+ I(\varepsilon_1^+ > c)]$ that

$$\begin{aligned} E\left[(\varepsilon_1 - \varepsilon_2)^+ I\left((\varepsilon_1 - \varepsilon_2)^+ > \frac{c}{2}\right)\right] &\geqslant E\left[\varepsilon_2^- I\left(\varepsilon_1 > \frac{c}{2}\right)\right] \\ &\geqslant E[\varepsilon_2^-] \cdot P(\varepsilon_1 > c/2) \\ &\geqslant \frac{1}{3} E\left[\varepsilon_1^+ I\left(\varepsilon_1^+ > c\right)\right]. \qquad \square \end{aligned}$$

Proof of Theorem 2.2. This proof follows the same pattern as that of Theorem 2.1.

(i) Suppose that, besides (A1), (A2)(i) is fulfilled. Since $\varphi_\varepsilon \in \Phi^{(i)}$ we have

$$\rho\left(\widehat{\widehat{\varphi}}_{X,n}, \widehat{\widehat{\varphi}}_{\varepsilon,n}; \widehat{\varphi}_{Z,n}\right) \leqslant \rho\left(\varphi_X, \varphi_\varepsilon; \widehat{\varphi}_{Z,n}\right) + \delta_n,$$

which yields, in conjunction with (3.5), that

$$\begin{aligned}\rho\left(\widehat{\widehat{\varphi}}_{X,n}, \widehat{\widehat{\varphi}}_{\varepsilon,n}; \varphi_Z\right) &\leqslant \rho\left(\widehat{\widehat{\varphi}}_{X,n}, \widehat{\widehat{\varphi}}_{\varepsilon,n}; \widehat{\varphi}_{Z,n}\right) + \rho\left(\varphi_X, \varphi_\varepsilon; \widehat{\varphi}_{Z,n}\right) \\ &\leqslant 2\rho\left(\varphi_X, \varphi_\varepsilon; \widehat{\varphi}_{Z,n}\right) + \delta_n \xrightarrow{\text{a.s}} 0. \end{aligned} \tag{3.16}$$

We fix an arbitrary elementary event of the underlying probability space such that the convergence in (3.16) takes place.

Analogously to the proof of Theorem 2.1, we can conclude from (3.16) that any subsequence $(n_k)_{k\in\mathbb{N}}$ of $\mathbb{N}$ contains a further subsequence $(n_k')_{k\in\mathbb{N}}$ such that, as $k \to \infty$,

$$\widehat{\widehat{P}}_{X,n_k'} \implies P_{X+c} \quad \text{and} \quad \widehat{\widehat{P}}_{\varepsilon,n_k'} \implies P_{\varepsilon-c},$$

for some $c \in \mathbb{R}$. The latter convergence implies in particular that $P_{\varepsilon-c}((-\infty,0]) \geqslant \limsup_{k\to\infty} \widehat{\widehat{P}}_{\varepsilon,n_k'}((-\infty,0]) \geqslant \frac{1}{2}$ and $P_{\varepsilon-c}([0,\infty)) \geqslant \limsup_{k\to\infty} \widehat{\widehat{P}}_{\varepsilon,n_k'}([0,\infty)) \geqslant \frac{1}{2}$, that is, 0 is a median of $P_{\varepsilon-c}$. Since, by assumption, 0 is the unique median of P_ε we conclude that $c = 0$. Since the above properties hold for a suitable subsequence $(n_k')_{k\in\mathbb{N}}$ of any arbitrary sequence $(n_k)_{k\in\mathbb{N}}$, they hold also for the full sequence. Hence, we have that

$$\widehat{\widehat{P}}_{X,n} \implies P_X \quad \text{and} \quad \widehat{\widehat{P}}_{\varepsilon,n} \implies P_\varepsilon,$$

for all elementary events such that the convergence in (3.16) takes place, hence, with probability 1.

(ii) Now we suppose that instead of (A2)(i) the alternative condition (A2)(ii) is fulfilled. First, it follows from the strong law of large numbers that

$$\widehat{\mu}_{1,n} \xrightarrow{\text{a.s}} \mu_1. \tag{3.17}$$

Therefore, there exists a sequence $(\widetilde{\varphi}_{\varepsilon,n})_{n\in\mathbb{N}}$ such that $\widetilde{\varphi}_{\varepsilon,n} \in \Phi_n^{(ii)}$ $\forall n \in \mathbb{N}$ and

$$\rho\left(\varphi_X, \widetilde{\varphi}_{\varepsilon,n}; \varphi_Z\right) \xrightarrow{\text{a.s}} 0,$$

which implies, in conjunction with (3.5), that

$$\rho\left(\varphi_X, \widetilde{\varphi}_{\varepsilon,n}; \widehat{\varphi}_{Z,n}\right) \leqslant \rho\left(\varphi_X, \widetilde{\varphi}_{\varepsilon,n}; \varphi_Z\right) + \rho\left(\varphi_X, \varphi_\varepsilon; \widehat{\varphi}_{Z,n}\right) \xrightarrow{\text{a.s.}} 0.$$

Taking into account that

$$\rho\left(\widehat{\widehat{\varphi}}_{X,n}, \widehat{\widehat{\varphi}}_{\varepsilon,n}; \widetilde{\varphi}_{Z,n}\right) \leqslant \rho\left(\varphi_X, \widetilde{\varphi}_{\varepsilon,n}; \widehat{\varphi}_{Z,n}\right) + \delta_n$$

we obtain

$$\begin{aligned}\rho\left(\widehat{\widehat{\varphi}}_{X,n}, \widehat{\widehat{\varphi}}_{\varepsilon,n}; \varphi_Z\right) &\leqslant \rho\left(\widehat{\widehat{\varphi}}_{X,n}, \widehat{\widehat{\varphi}}_{\varepsilon,n}; \widehat{\varphi}_{Z,n}\right) + \rho\left(\varphi_X, \varphi_\varepsilon; \widehat{\varphi}_{Z,n}\right) \\ &\leqslant \rho\left(\varphi_X, \widetilde{\varphi}_{\varepsilon,n}; \widehat{\varphi}_{Z,n}\right) + \rho\left(\varphi_X, \varphi_\varepsilon; \widehat{\varphi}_{Z,n}\right) + \delta_n \xrightarrow{\text{a.s.}} 0. \end{aligned} \tag{3.18}$$

We fix an arbitrary elementary event of the underlying probability space such that the convergence in (3.17) and (3.18) takes place.

As in the proof of Theorem 2.1, we can conclude from (3.18) that any subsequence $(n_k)_{k\in\mathbb{N}}$ of $\mathbb{N}$ contains a further subsequence $(n'_k)_{k\in\mathbb{N}}$ such that, as $k \to \infty$,

$$\widehat{\widehat{P}}_{X,n'_k} \Longrightarrow P_{X+c} \quad \text{and} \quad \widehat{\widehat{P}}_{\varepsilon,n'_k} \Longrightarrow P_{\varepsilon-c}, \tag{3.19}$$

for some $c \in \mathbb{R}$. Let $\widehat{\widehat{\varepsilon}}_{n'_k,1}$ and $\widehat{\widehat{\varepsilon}}_{n'_k,2}$ be independent random variables with common distribution $\widehat{\widehat{P}}_{\varepsilon,n'_k}$. Since

$$\widehat{\widehat{\varepsilon}}_{n'_k,1} - \widehat{\widehat{\varepsilon}}_{n'_k,2} \xrightarrow{d} \varepsilon_{11} - \varepsilon_{12} \tag{3.20}$$

and

$$E_{\widehat{\widehat{P}}_{\varepsilon,n'_k}} |\widehat{\widehat{\varepsilon}}_{n'_k,1} - \widehat{\widehat{\varepsilon}}_{n'_k,2}| \leqslant \widehat{\mu}_{1,n} \underset{k\to\infty}{\longrightarrow} \mu_1 = E|\varepsilon_{11} - \varepsilon_{12}|$$

we conclude that

$$E_{\widehat{\widehat{P}}_{\varepsilon,n'_k}} |\widehat{\widehat{\varepsilon}}_{n'_k,1} - \widehat{\widehat{\varepsilon}}_{n'_k,2}| \underset{k\to\infty}{\longrightarrow} E|\varepsilon_{11} - \varepsilon_{12}|.$$

This implies, again in conjunction with (3.20), that the sequence $(\widehat{\widehat{\varepsilon}}_{n'_k,1} - \widehat{\widehat{\varepsilon}}_{n'_k,2})_{k\in\mathbb{N}}$ is uniformly integrable; see e.g. Chung [6, Theorem 4.5.4]. By Lemma 3.1, this implies that $(\widehat{\widehat{\varepsilon}}_{n'_k,1})_{k\in\mathbb{N}}$ is also uniformly integrable. Therefore, we conclude from (3.19) that

$$0 = E_{\widehat{\widehat{P}}_{\varepsilon,n'_k}} \widehat{\widehat{\varepsilon}}_{n'_k,1} \underset{k\to\infty}{\longrightarrow} E_{P_\varepsilon}\varepsilon - c,$$

which means that c must be zero. Hence, we have that

$$\widehat{\widehat{P}}_{X,n'_k} \Longrightarrow P_X \quad \text{and} \quad \widehat{\widehat{P}}_{\varepsilon,n'_k} \Longrightarrow P_\varepsilon.$$

This also holds for the full sequence, with probability 1. This completes the proof. □

Acknowledgment

I thank two referees for their valuable comments.

References

[1] J. Aczél, Lectures on Functional Equations and Their Applications, Academic Press, New York, 1966.
[2] P. Billingsley, Convergence of Probability Measures, Wiley, New York, 1968.
[3] C. Butucea, C. Matias, Minimax estimation of the noise level and of the deconvolution density in a semiparametric deconvolution model, Bernoulli 11 (2005) 309–340.
[4] R.J. Carroll, P. Hall, Optimal rates of convergence for deconvolving a density, J. Amer. Statist. Assoc. 83 (1988) 1184–1186.
[5] L. Cavalier, N.W. Hengartner, Adaptive estimation for inverse problems with noisy operators, Inverse Problems 21 (2005) 1345–1361.
[6] K.L. Chung, A Course in Probability Theory, second ed., Academic Press, San Diego, 1974.
[7] L. Devroye, Consistent deconvolution in density estimation, Canad. J. Statist. 17 (1989) 235–239.
[8] P. Diggle, P. Hall, A Fourier approach to nonparametric deconvolution of a density estimate, J. Roy. Statist. Soc. Ser. B 55 (1993) 523–531.
[9] D.L. Donoho, R.C. Liu. Geometrizing rates of convergence, I, Technical Report 137a, Department of Statistics, University of California, Berkeley, 1987.

[10] S. Efromovich, Density estimation for the case of supersmooth measurement error, J. Amer. Statist. Assoc. 92 (1997) 526–535.
[11] S. Efromovich, V. Koltchinskii, On inverse problems with unknown operators, IEEE Trans. Inform. Theory 47 (2001) 2876–2893.
[12] J. Fan, On the optimal rates of convergence for nonparametric deconvolution problems, Ann. Statist. 19 (1991) 1257–1272.
[13] J. Fan, Global behavior of deconvolution kernel estimates, Statist. Sinica 1 (1991) 541–551.
[14] J. Fan, Adaptively local one-dimensional subproblems with application to a deconvolution problem, Ann. Statist. 21 (1993) 600–610.
[15] W. Feller, An Introduction to Probability Theory and its Applications, Wiley, New York, 1966.
[16] W.A. Fuller, Measurement Error Models, Wiley, New York, 1987.
[17] P. Hall, Q. Yao, Inference in components of variance models with low replication, Ann. Statist. 31 (2003) 414–441.
[18] J.L. Horowitz, M. Markatou, Semiparametric estimation of regression models for panel data, Rev. Econom. Studies 63 (1996) 145–168.
[19] A.I. Il'inskii, On the zeros, the argument of a characteristic function, Teor. Veroyatn. Primen. 20 (1975) 421–427 (in Russian) (English translation: Theory Probab. Appl. 20, 410–415).
[20] I. Kotlarski, On characterizing the gamma and the normal distribution, Pacific J. Math. 20 (1967) 69–76.
[21] T. Li, Q. Vuong, Nonparametric estimation of the measurement error model using multiple indicators, J. Multivariate Anal. 65 (1998) 139–165.
[22] M.C. Liu, R.C. Taylor, A consistent nonparametric density estimator for the deconvolution problem, Canad. J. Statist. 17 (1990) 427–438.
[23] A. Meister, On the effect of misspecifying the error density in a deconvolution problem, Canad. J. Statist. 32 (2004) 439–449.
[24] A. Meister, Deconvolution density estimation with a testing procedure for the error distribution. University of Stuttgart, Preprint, 2004.
[25] A. Meister, Density estimation with normal measurement error with unknown variance, Statist. Sinica 16 (2006) 195–211.
[26] J. Mendelsohn, J. Rice, Deconvolution of microfluorometric histograms with *B* splines, J. Amer. Statist. Assoc. 77 (1982) 748–753.
[27] M.H. Neumann, On the effect of estimating the error density in nonparametric deconvolution, J. Nonparametric Statist. 7 (1997) 307–330.
[28] F.H. Ruymgaart, A unified approach to inversion problems in statistics, Math. Methods Statist. 2 (1993) 130–146.
[29] L.A. Stefanski, R.J. Carroll, Deconvoluting kernel density estimators, Statistics 21 (1990) 169–184.
[30] L.A. Stefanski, R.J. Carroll, Deconvolution-based score tests in measurement error models, Ann. Statist. 19 (1991) 249–259.
[31] N.G. Ushakov, Selected Topics in Characteristic Functions, VSP, Utrecht, 1999.